SELF-EVALUATION FOR VIDEO TRACKING SYSTEMS

Hao Wu and Qinfen Zheng
Centre for Automation Research
Dept. of Electrical and Computer Engineering
University of Maryland, College Park, MD-20742
{wh2003, qinfen}@cfar.umd.edu

ABSTRACT

In this paper, we present an algorithm for automatic performance evaluation of a video tracking system that does not require ground-truth data. Such an algorithm can play an important role in automatically determining when the underlying system loses track and needs re-initialization. The algorithm is based on measuring appearance similarity and tracking uncertainty. Several experimental results on vehicle and human tracking are presented. Effectiveness of the evaluation scheme is assessed by comparisons with ground truth. The proposed self evaluation algorithm has been used in an acoustic/video based moving vehicle detection and tracking system.

1. INTRODUCTION

An object tracking system can fail under many circumstances. It could be due to illumination changes, pose variation, occlusion, and other factors. There is a need for automatic performance evaluation. Most of the existing work on tracking performance evaluation has focused on overall algorithmic performance evaluation using ground-truth data. Their usefulness in real time determining tracking failure is quite limited. In this paper, we present a tracker self-evaluation algorithm that automatically evaluates the tracking quality on-the-run and does not require ground-truth data.

Online self-evaluation for keeping track of system performance has been studied for video based object segmentation. In [Erdem, 2004], segmentation and motion consistency along the object contour and histogram similarity are calculated and used to evaluate the goodness of segmentation and tracking. However, a generic tracking algorithm may not segment the object from the background and hence, the contour information may not be available. We address video tracking systems whose targets are bounded by boxes. The track assessment is mainly based on appearance similarity and trajectory smoothness. We reduce the confidence in tracking when there is ambiguity in the result. The uncertainty is assessed through monitoring several ambiguity measurements.

The paper is organized as follows: ambiguity feature extraction and track evaluation criterion are discussed in Section 2 and 3 respectively; Section 4 gives several experimental results; finally conclusions are given in Section 5.

2. FEATURES USED FOR SELF-EVALUATION

In a common video tracker, the location and appearance of the target is represented through a representative chip specified by a bounding box in the image frame. Contour based trackers can be modified to fit into such a framework. Intuitively, one may think that the appearance change can be used for evaluation. However, it is not reliable to judge the tracking performance solely based on the appearance of the tracking box. Appearance change may be caused by two factors: (1) object pose change due to camera and/or object motion and (2) appearance difference measure not consistent with subjective evaluation. The appearance change doesn't necessarily indicate poor tracking performance. In addition, in many cases the bounding box includes some background pixels, which makes the appearance evaluation difficult.

In our experience on video surveillance using static infrared camera, we have noticed that when tracking fails, the size and location of bounding box changes irregularly. Once the tracking bounding box locks onto background pixels, it changes randomly due to the similarity of the background clutter. Another common cause of tracking failure is that the tracking bounding box locks onto background objects. Our goal is to detect any tracking failure soon after it occurs. The following ambiguity tests are examined in our self evaluation algorithm.

Test 1: Trajectory complexity evaluation

Normally, a moving vehicle will not change its direction and speed dramatically in a few adjacent frames. Therefore, rapid and frequent change in object motion trajectory is a sign of tracking failure. We measure trajectory complexity as the ratio of the trajectory path

[1] Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-02-0008.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **00 DEC 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Self-Evaluation For Video Tracking Systems**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Centre for Automation Research Dept. of Electrical and Computer Engineering University of Maryland, College Park, MD-20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001736, Proceedings for the Army Science Conference (24th) Held on 29 November - 2 December 2004 in Orlando, Florida., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

length, $L_{p_1p_2}$, and end points distance, $D_{p_1p_2}$, between two tracking points $p_1 = p_{(t-t)}$ and $p_2 = p_{(t)}$ as shown in Fig. 1. Normally, the larger the ratio is, the more complex the trajectory will be. We define trajectory complexity indicator as

$$I_1(t) = \begin{cases} 0 & \text{if } \dfrac{L_{p_1p_2}}{D_{p_1p_2}} \geq T_1 \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

We can further include trajectory direction change in trajectory complexity indicator.

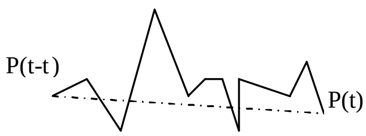


Fig.1 Illustration of tracking trajectory

Test 2: Motion smoothness evaluation

We noticed that the trajectory increment between two adjacent frames often increases when tracking fails. We define motion step as the displacement of object box over two consecutive frames, $D_{p_{(t-1)}p_{(t)}}$. Motion smoothness indicator is defined as

$$I_2(t) = \begin{cases} 0 & \text{if } D_{p_{(t-1)}p_{(t)}} \geq T_2 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

The threshold T_2 is determined according to prior knowledge of object motion. For object tracking from a moving camera, camera ego motion should first be estimated and removed from the object displacement computation.

Test 3: Scale constancy evaluation

In general, for medium to long range surveillance, we expect the scale change to be small. We measure target scale change as the ratio of the area of current target bounding box, A_t , to the area of initial bounding box, A_0 . Both the target scale change and scale change rate are measured and used in track evaluation. We define the scale constancy indicator as

$$I_3(t) = \begin{cases} 0 & \text{if } \left\{ \begin{array}{l} \left\{ \dfrac{A_t}{A_0} \leq T_{31} \right\} \cup \left\{ \dfrac{A_t}{A_0} \geq T_{32} \right\} \cup \\ \left\{ \dfrac{dA_t}{A_0 dt} \leq T_{33} \right\} \cup \left\{ \dfrac{dA_t}{A_0 dt} \geq T_{34} \right\} \end{array} \right\} \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

Test 4: Shape similarity evaluation

Shape is an important discriminator for objects. When the tracking bounding box switches to a different object or to the background, the shape of the bounding box often also changes. We use aspect ratio, $\mathrm{Width}/\mathrm{Height}$, of the bounding box to represent object shape and measure the shape similarity as the ratio of bounding box aspect ratios. The shape similarity indicator is defined as

$$I_4(t) = \begin{cases} 0 & \text{if } \left\{ \dfrac{\mathrm{W_t/H_t}}{\mathrm{W_0/H_0}} \leq T_{41} \right\} \cup \left\{ \dfrac{\mathrm{W_t/H_t}}{\mathrm{W_0/H_0}} \geq T_{42} \right\} \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Test 5: Appearance similarity evaluation

Although tracking evaluation should not solely depend on appearance similarity, appearance change often results in tracking failure. Therefore, quantifying the appearance change is still important. We use three appearance change measures to evaluate the appearance stability. The first one, D_I , is pixel by pixel difference between the current object and the initial object; the second one, D_H , is difference of image intensity histograms between the current and initial objects as used in 0; the third one, D_M , is the sum of weighted differences between the current appearance model and the initial appearance model. Other measurement methods can also be added. We define the appearance similarity indicator as

$$I_5(t) = \begin{cases} 0 & \text{if } \left\{ \mathrm{D_I} \geq T_{51} \right\} \cup \left\{ \mathrm{D_H} \geq T_{52} \right\} \cup \left\{ \mathrm{D_M} \geq T_{53} \right\} \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

3. EVALUATION CRITERION

In ideal situation, a good tracking should have all the five tracking evaluation indicators equal to one. In practical circumstances, some unexpected factors may trigger one or two of these indicators, while the tracking performance is still good. However if three or more indicators have been triggered, we conclude that the tracking performance has deteriorated. We fuse the above five test scores to get a comprehensive tracking performance score. We first learn the uncertainty decision thresholds for each test using empirical data and then compute a weighted sum of the five indicators

$$q(t) = \sum_{i=1}^{5} w_i I_i(t), \qquad \sum_{i=1}^{5} w_i = 1 \quad (6)$$

In general, the larger the $q(t)$ is, the better the tracking performance. When $q(t)$ drops below a threshold, we conclude that the tracking performance has deteriorated and needs to be re-initialized. The weight can be learnt from training data. In our implementation, the appearance weight, w_5, is set slightly larger than others. In implementation, one may re-initialize the system only after $q(t)$ is below a threshold for a specified period of time.

4. EXPERIMEN RESULTS

The proposed algorithm was tested on different surveillance videos. Fig.2 shows evaluation results on an IR vehicle surveillance sequence. The vehicle first moved straight away from the camera and then made a left turn. The results show that the self evaluation algorithm does give a good indication of the tracking performance. In Fig. 2(a), when the bounding box does not fit the object well, the evaluation score drops. After re-initialization, the bounding box fits the object and the evaluation score rises, as shown in Fig. 2(b). We also compared the self evaluation result with ground truth (Fig.3). It is shown that as the distance between the tracked object location and the ground truth increases, our tracking confidence score decrease indicating deterioration in tracking performance. When integrated into a moving vehicle detection and tracking system [Sankaranayanan, 2004], the proposed algorithm helps the video surveillance system maintaining a good target track by re-initializing the tracker whenever the tracker performance deteriorates. The tracking algorithm used in our experiments is the adaptive appearance model based tracker developed by Zhou, et al [Zhou, 2004].

Fig. 4 shows the results of evaluation of pedestrian detection and tracking from a color surveillance video. The first three images are representative frames of the surveillance video with the tracking bounding box superimposed. The corresponding tracker evaluation scores are shown in the bottom row of Fig.4. In this example, the bounding box switches to the background and wanders around at that position afterwards. Our self evaluation criterion correctly reports the tracking failure.

Fig.5 shows the results of evaluating a pedestrian tracking with partial occlusion and reappearance. The tracked person walks behind a moving car. The tracker becomes uncertain while partially occluded by the moving vehicle. The tracker regains its confidence/performance after the human reappears. Our tracker evaluation algorithm correctly scores the event.

Fig.6 shows the evaluation results for tracking a group of pedestrian with significant occlusion. As the tracked human group is blocked by the moving van, the bounding box switches to the van and loses the target. Our self-evaluation score drops when the tracker fails. We expect the confidence score will drop further if target trajectory direction is also incorporated in the evaluation measurements.

5. CONCLUSIONS

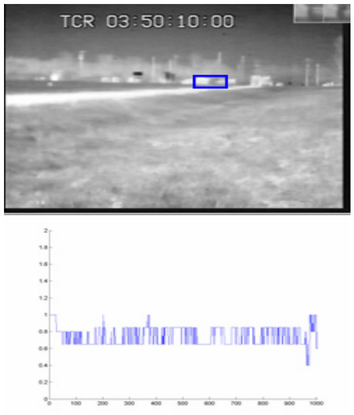
In this paper, we present an algorithm for automatic performance evaluation of a video tracking system that does not require ground-truth data. The algorithm is based on measuring appearance similarity and tracking uncertainty. Several experimental results on vehicle and human tracking are reported. Effectiveness of the evaluation scheme is demonstrated by comparisons with ground truth. The proposed self evaluation algorithm has been used in an acoustic/video based moving vehicle detection and tracking system [Sankaranayanan, 2004].

6. REFERENCES

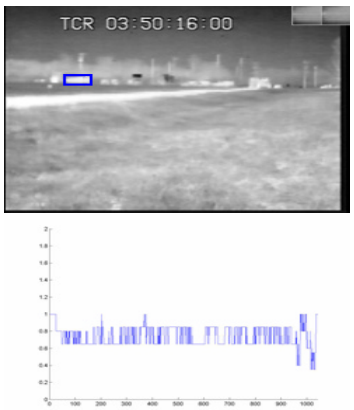
Erdem, C.E. Sankur, B, Tekalp, A.M., 2004: Performance Measures for Video Object Segmentation and Tracking, *IEEE Trans. Image Processing,* **13**:931-951.

Sankaranayanan, A.C., et al, 2004: Vehicle Tracking using Acoustic and Video Sensors, *Proc. 24th Army Science Conference* (to appear).

Zhou, S., Chellappa, R., Moghaddam, B., 2004: Visual Tracking and Recognition Using Appearance-adaptive Models in Particle Filters, *IEEE Trans. Image Processing* (to appear).



(a)



(b)

Fig.2 Improved video tracking with track evaluation and appearance updating. Also shown are the corresponding evaluation plots.

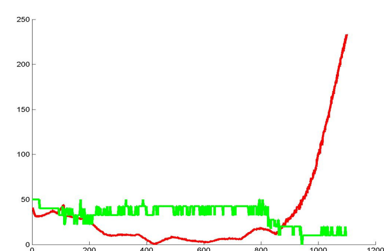


Fig.3 Comparison of self-evaluation score and the ground truth. The red line is the distance between GPS measurements and tracked target center; the green line is the evaluation scores reported by our algorithm.

Fig.4 An example of pedestrian tracking. Shown in the top three rows are representative frames with the tracking bounding box superimposed. The corresponding tracker evaluation scores are shown in the bottom row. Our self evaluation criterion correctly reports the tracking failure.

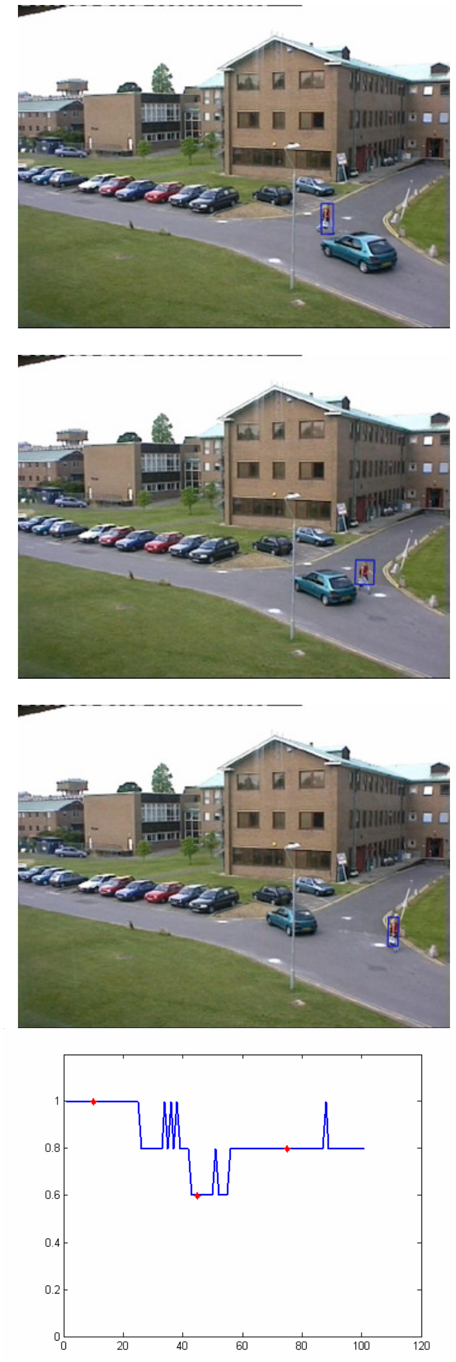


Fig.5. An example of tracking pedestrian with partial occlusion. The tracked person walks behind a moving car. The tracker becomes uncertain while partially occluded by the moving vehicle. The tracker regains its performance after the human is cleared of occlusion. Our tracker evaluation algorithm correctly scores the event.

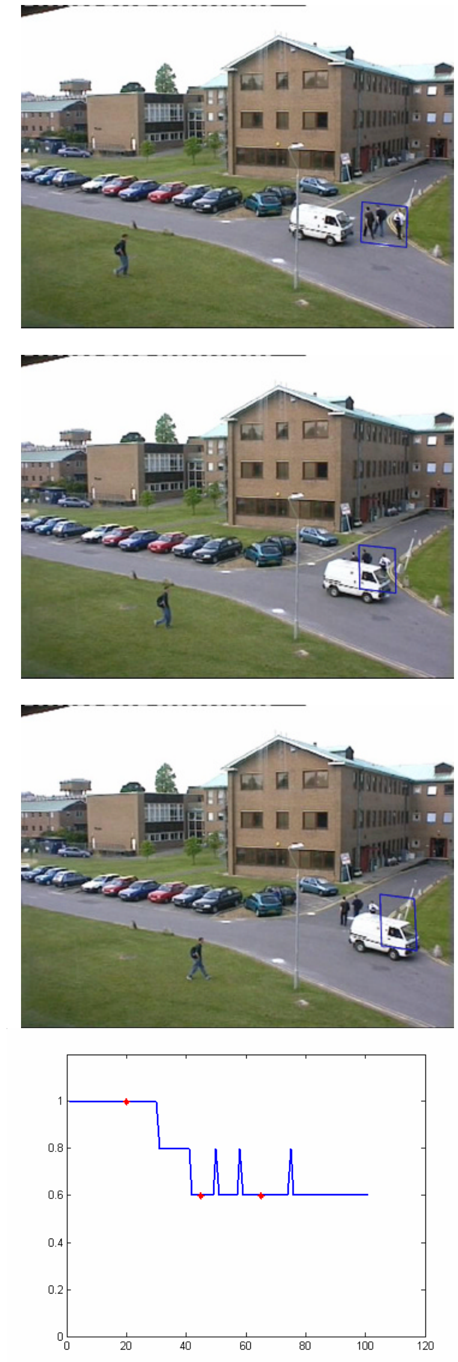


Fig.6. An example of tracking a group of pedestrian with significant occlusion. As the tracked human group is occluded by the moving van, the bounding box switches to the van and lose the target. Our self-evaluation score drops when the tracker fails.